



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **07140999 A**(43) Date of publication of application: **02 . 06 . 95**

(51) Int. Cl.

G10L 5/04
G10L 3/00
(21) Application number: **05308731**(22) Date of filing: **15 . 11 . 93**(71) Applicant: **SONY CORP**(72) Inventor: **OIKAWA YOSHIAKI**
YAMADA KEIICHI(54) **DEVICE AND METHOD FOR VOICE SYNTHESIS**

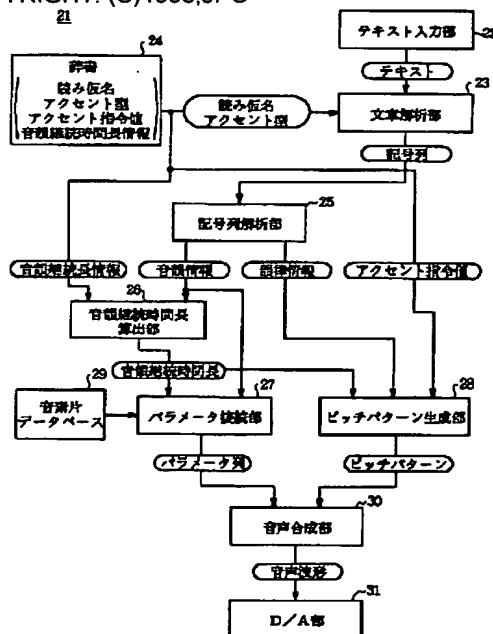
(57) Abstract:

PURPOSE: To output a synthesized voice which is close to human uttering by generating a parameter column employing the continuation time length of phonemes, generating a pitch pattern based on the phoneme continuation time length and phoneme information and synthesizing voice waveforms by the parameter column and the pitch pattern.

CONSTITUTION: A phoneme continuation time length information is stored in advance in a dictionary 24 and the continuation time length of each phoneme is computed. The computed continuation time lengths are outputted to a parameter connection section 27 and a pitch pattern generating section 28. In the section 27, the phoneme piece data selected from a phoneme piece database 29 are held based on the phoneme information from a symbol column analysis section 25 and the continuation time length computed by a phoneme continuation time length computing section 26 and a parameter column is generated. On the other hand, a pitch pattern is generated in a pitch pattern generating section 28 based on the accent specification values from the phoneme information and the dictionary 24 and the continuation time length. In a voice synthesis section 30, synthesized voice signals are obtained based on the

parameter columns and the pitch patterns.

COPYRIGHT: (C)1995,JPO



Best Available Copy

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平7-140999

(43)公開日 平成7年(1995)6月2日

(51)Int.Cl.⁶

G 1 0 L 5/04
3/00

識別記号

F
H

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数2 F D (全 8 頁)

(21)出願番号 特願平5-308731

(22)出願日 平成5年(1993)11月15日

(71)出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72)発明者 及川 芳明

東京都品川区北品川6丁目7番35号ソニー株式会社内

(72)発明者 山田 敬一

東京都品川区北品川6丁目7番35号ソニー株式会社内

(74)代理人 弁理士 田辺 恵基

(54)【発明の名称】 音声合成装置及び音声合成方法

(57)【要約】

【目的】本発明は、人間の発声に近い合成音声を生成することができる音声合成装置及び音声合成方法を実現する。

【構成】読み仮名、アクセント型等の情報を保持するメモリ手段24にアクセント指令値及び又は音韻継続時間長情報を予め用意する。音韻情報又は音韻情報及び音韻継続時間長情報に基づいて算出した音韻の継続時間長を用いて音素片データのパラメータ列を生成する。また算出した音韻の継続時間長及び韻律情報又は算出した音韻の継続時間長、韻律情報及びアクセント指令値に基づいてピッチパターンを生成する。このようにして生成されたパラメータ列及びピッチパターンにより音声波形を合成することにより、人間の発声に一段と近い合成音声出力することができる。

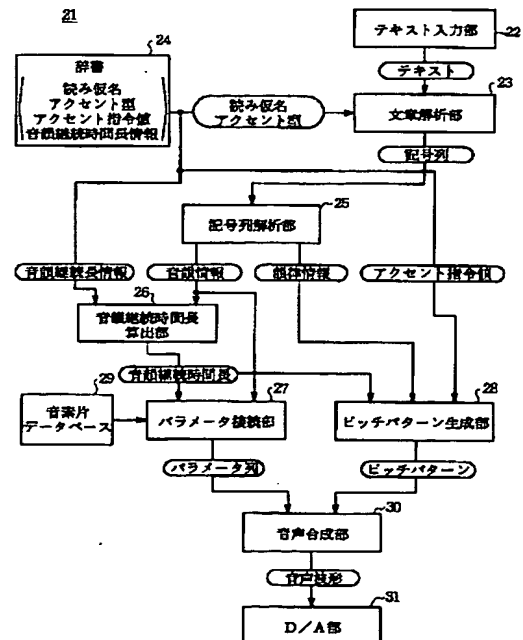


図1 実施例における音声合成装置の構成

【特許請求の範囲】

【請求項 1】読み仮名、アクセント型等の情報の他に音韻継続時間長情報及び又はアクセント指令値を保持するメモリ手段を基準にして、入力された文章を解析して当該文章の音韻情報及び韻律情報を示す記号列データを出力する文章解析手段と、

上記記号列データを音韻情報及び韻律情報に分離して抽出する記号列解析手段と、

上記音韻情報又は上記音韻情報及び上記メモリ手段に保持されている音韻継続時間長情報に基づいて音韻の継続時間長を算出する音韻継続時間長算出手段と、

上記音韻情報及び上記音韻の継続時間長に基づいてパラメータ列を生成するパラメータ接続手段と、

上記音韻の継続時間長及び上記韻律情報又は上記音韻の継続時間長、上記韻律情報及び上記メモリ手段に保持されているアクセント指令値に基づいてピッチパターンを生成するピッチパターン生成手段と、

上記パラメータ列及び上記ピッチパターンより音声波形を合成する音声合成手段とを具えることを特徴とする音声合成装置。

【請求項 2】読み仮名、アクセント型等の情報の他に音韻継続時間長情報及び又はアクセント指令値を保持するメモリ手段を基準にして、入力された文章を解析して当該文章の音韻情報及び韻律情報を示す記号列データを出力し、

上記記号列データを音韻情報及び韻律情報に分離して抽出し、

上記音韻情報又は上記音韻情報及び上記メモリ手段に保持されている音韻継続時間長情報に基づいて音韻の継続時間長を算出し、

上記音韻情報及び上記音韻の継続時間長に基づいてパラメータ列を生成し、

上記音韻の継続時間長及び上記韻律情報又は上記音韻の継続時間長、上記韻律情報及び上記メモリ手段に保持されているアクセント指令値に基づいてピッチパターンを生成し、

上記パラメータ列及び上記ピッチパターンより音声波形を合成することを特徴とする音声合成方法。

【発明の詳細な説明】

【0001】

【目次】以下の順序で本発明を説明する。

産業上の利用分野

従来の技術（図 4）

*

$$\text{母音継続時間長} = \text{母音平均継続時間長} + \sum_j \sum_k^j a(j, k) \sigma(j, k)$$

..... (1)

のように母音の平均継続時間長と、音韻環境の各要因における母音平均長からの変動分の時間長とを加えることにより求める手法がある。この手法は平成 2 年 3 月の日

* 発明が解決しようとする課題

課題を解決するための手段（図 1）

作用（図 1）

実施例（図 1～図 3）

発明の効果

【0002】

【産業上の利用分野】本発明は音声合成装置及び音声合成方法に関し、特に入力文字系列より音声を合成するテキスト音声合成装置及びテキスト音声合成方法に適用して好適なものである。

【0003】

【従来の技術】従来、音声合成装置においては一般に図 4 に示すような処理を行なうことにより音声合成を行なっている。まず音声合成装置 1 において、所定の入力装置よりテキスト入力部 2 に漢字仮名混じりの日本語テキストが入力され、テキストを文章解析部 3 に出力する。文章解析部 3 は入力されたテキストを辞書 4 を基準にして解析し、読み仮名文字列に変換した後、単語固有のアクセント型を検出して単語、文節毎に分解する。

【0004】すなわち日本語においては、英語のように単語が分かち書きされていないことから、例えば「米国産業界」のような言葉は、「米国／産業・界」、「米／国産／業界」のように 2 種類に区分化することができ

る。【0005】このため文章解析部 3 は辞書 4 を参考にしながら、言葉の連続関係及び単語の統計的性質を利用して、テキスト入力を単語、文節毎に分解するようになされ、これにより単語、文節の境界を検出する。さらに文章解析部 3 は、このようにして求められた各文章毎の読み仮名（音韻に関する情報）と文節境界及びアクセント型（韻律に関する情報）とを示す記号列を記号列解析部 5 に出力する。

【0006】記号列解析部 5 はこの記号列を音韻情報及び韻律情報に分離して抽出し、音韻情報は音韻継続時間長算出部 6 及びパラメータ接続部 7 に出力され、韻律情報はピッチパターン生成部 8 に出力される。ここで音韻情報とは発声される音に関する情報、韻律情報とはアクセントやイントネーションに関する情報である。

【0007】音韻継続時間長算出部 6 に音韻情報が入力されると、音韻継続時間長算出部 6 は音韻情報より各音韻の継続時間長を算出する。例えば音韻の継続時間長を算出する方法として、母音の継続時間長を次式

【数 1】

$$\text{母音継続時間長} = \text{母音平均継続時間長} + \sum_j \sum_k^j a(j, k) \sigma(j, k)$$

..... (1)

本音響学会講演論文集（海木他著）に発表されている。

【0008】この（1）式において、 a は係数行列、 σ は母音の音韻環境が各要因のどのカテゴリに該当するか

を示す係数「1」又は「0」の行列、 m は要因数（すなわち母音の種類、前方音韻の種類、前々方音韻の種類、後方音韻の種類、後々方音韻の種類、前方が促音、後方が促音、長音、呼吸段落モーラ数、呼吸段落内位置、文モーラ数、文内位置）、 l は各要因のカテゴリ数を表す。

【0009】この場合、入力として与えられるパラメータは求めようとしている音韻の音韻環境であり、母音平均長からの変動分は予め大量の文章を分析して求めておくことにより実現される。

【0010】このようにして求められた音韻の継続時間長はパラメータ接続部7及びピッチパターン生成部8に*

$$l_n(F_n(t)) = l_n(F_{n,n}) + \sum_i A_{n,i} G_{n,i}(t - T_{1,i})$$

$$+ \sum_i A_{n,i} [G_{n,i}(t - T_{1,i}) - G_{n,i}(t - T_{2,i})]$$

..... (2)

に示すようにピッチパターンをフレーズ成分とアクセント成分とに分け、それぞれがインパルス入力、ステップ入力で駆動される2次臨界応答出力の和で表現するモデルがある。このモデルは1989年1月の通信学会論文誌 Vol. J72-A, No. 1 (藤崎他著「基本周波数パターン生成過程モデルに基づく文章音声の合成」) に発表されている *

$$G_{n,i}(t) = \alpha^2 \cdot t \exp(-\alpha \cdot t)$$

及び次式

$$G_{n,i}(t) = \min [1 - (1 + \beta \cdot t) \exp(-\beta \cdot t), \theta]$$

となる。

【0014】この(3)式及び(4)式において、 F_{min} は最低ピッチ周波数、 α 、 β 、 θ は定数、 A_{pi} はフレーズ指令の大きさ、 $T_{1,i}$ はフレーズ指令の位置、 A_{aj} はアクセント指令の大きさ、 $T_{1,j}$ はアクセント指令の開始位置、 $T_{2,j}$ はアクセント指令の終了位置である。また $t < 0$ では $G_{pi}(t) = G_{aj}(t) = 0$ となる。

【0015】この場合、入力として与えられるパラメータは、入力の時刻、大きさ及び応答の速さを決める係数である。これらの値は、予め大量の文章を分析して求めて保持しておき、所定の規則により適切な値を用いてピッチパターンを生成する。次に音声合成部10はパラメータ列及びピッチパターンに基づいて波形合成処理を行い、デジタルアナログ変換部(D/A部)11を介して合成音声信号を出力する。

【0016】

【発明が解決しようとする課題】ところでこのような従来のテキスト音声合成装置1では、ピッチパターンの算出は、テキスト解析のための辞書に予め記載されている各単語のアクセント型と、単語が連なつた場合のアクセントの変化規則とにより、文節のアクセントが求められて記号列に出力されている。この場合、同じアクセント型で同じモーラ数の単語が文章中の同じ位置に出現した

*出力される。パラメータ接続部7では、音韻情報及び算出した各音韻の継続時間長に基づいて、音素片データベース9より読み出した音素片データを接続してパラメータ列を生成する。生成されたパラメータ列は音声合成部10に出力される。

【0011】他方、ピッチパターン生成部8では、韻律情報及び算出した各音韻の継続時間長に基づいてピッチパターンを生成し、生成されたピッチパターンは音声合成部10に出力される。

10 【0012】ここでピッチパターン算出の例として、次式

【数2】

※る。

20 【0013】この(2)式において、 $G_{pi}(t)$ 、 $G_{aj}(t)$ はそれぞれフレーズ制御機構のインパルス応答、アクセント制御機構のステップ応答であり、 $t \geq 0$ の範囲ではインパルス応答、ステップ応答はそれぞれ次式

【数3】

..... (3)

【数4】

..... (4)

場合には、求められるアクセントパターンは同じものとなる。

30 【0017】また音韻の継続時間長は、算出しようとしている音韻の前後、その前後の音韻環境や文章中の位置を考慮して求められるが、同じ音韻環境で異なる単語の場合には、求められる音韻の継続時間長は同じものとなる。

40 【0018】ところが実際に人間が発声した場合のピッチのパターンは、同じアクセント型で同じモーラ数の単語が文章中の同じ位置に出現しても、同じピッチパターンで話されることはほとんどなく、音韻の継続時間長もミクロ的に同じ音韻環境であつても単語単位のようなマクロの音韻環境を考えると同様の長さになるということは少ない。従つて従来のテキスト音声合成装置の合成音声は、単調になつてしまうという問題があつた。

【0019】本発明は以上の点を考慮してなされたもので、人間の発声に近い合成音声を得ることができる音声合成装置及び音声合成方法を提案しようとするものである。

【0020】

50 【課題を解決するための手段】かかる課題を解決するため本発明においては、読み仮名、アクセント型等の情報の他に音韻継続時間長情報及び又はアクセント指令値を

保持するメモリ手段 24 を基準にして、入力された文章を解析して当該文章の音韻情報及び韻律情報を示す記号列データを出力する文章解析手段 23 と、記号列データを音韻情報及び韻律情報に分離して抽出する記号列解析手段 25 と、音韻情報又は音韻情報及びメモリ手段 24 に保持されている音韻継続時間長情報に基づいて音韻の継続時間長を算出する音韻継続時間長算出手段 26 と、音韻情報及び音韻の継続時間長に基づいてパラメータ列を生成するパラメータ接続手段 27 と、音韻の継続時間長及び韻律情報又は音韻の継続時間長、韻律情報及びメモリ手段 24 に保持されているアクセント指令値に基づいてピッチパターンを生成するピッチパターン生成手段 28 と、パラメータ列及びピッチパターンより音声波形を合成する音声合成手段 30 とを設けるようにする。

【0021】また本発明においては、読み仮名、アクセント型等の情報の他に音韻継続時間長情報及び又はアクセント指令値を保持するメモリ手段 24 を基準にして、入力された文章を解析して当該文章の音韻情報及び韻律情報を示す記号列データを出力し、記号列データを音韻情報及び韻律情報に分離して抽出し、音韻情報又は音韻情報及びメモリ手段 24 に保持されている音韻継続時間長情報に基づいて音韻の継続時間長を算出し、音韻情報及び音韻の継続時間長に基づいてパラメータ列を生成し、音韻の継続時間長及び韻律情報又は音韻の継続時間長、韻律情報及びメモリ手段 24 に保持されているアクセント指令値に基づいてピッチパターンを生成し、パラメータ列及びピッチパターンより音声波形を合成するようにする。

【0022】

【作用】読み仮名、アクセント型等の情報を保持するメモリ手段 24 にアクセント指令値及び又は音韻継続時間長情報を予め用意する。パラメータ接続手段 27 では、音韻情報又は音韻情報及び音韻継続時間長情報に基づいて算出した音韻の継続時間長を用いて音素片データのパラメータ列を生成する。ピッチパターン生成手段 28 では、算出した音韻の継続時間長及び韻律情報又は算出した音韻の継続時間長、韻律情報及びアクセント指令値に基づいてピッチパターンを生成する。このように生成されたパラメータ列及びピッチパターンより音声波形を合成することにより、人間の発声に一段と近い合成音声

【0023】

【実施例】以下図面について、本発明の一実施例を詳述する。

【0024】図 1 において、21 は全体として音声合成装置を示し、テキスト入力部 22 に入力されたテキストを所定の各処理部で処理することにより人間の発声に近い音声

処理部の処理内容は異なる。

【0025】まず所定の入力装置よりテキスト入力部 22 に漢字仮名混じりの日本語テキストが入力され、これを文章解析部 23 に出力する。文章解析部 23 は入力されたテキストを辞書 24 を基準にして解析し、読み仮名文字列に変換した後、単語固有のアクセント型を検出して単語、文節毎に分解する。

【0026】すなわち文章解析部 23 は辞書 24 を参考にしながら、言葉の連続関係及び単語の統計的性質を利用して、テキスト入力を単語、文節毎に分解するようになされ、これにより単語、文節の境界を検出する。さらに文章解析部 23 は、このようにして求められた各文章毎の読み仮名（音韻に関する情報）と文節境界及びアクセント型（韻律に関する情報）とを示す記号列を記号列解析部 25 に出力する。

【0027】ここで辞書 24 には、アクセント型を示す情報の他に単語固有のアクセント指令値及び音韻継続時間長情報が予め保持され、音韻継続時間長算出部 26 は音韻継続時間長情報を読み出し、ピッチパターン生成部 27 はアクセント指令値を讀出す。次に記号列解析部 25 はこの記号列を音韻情報及び韻律情報に分離して抽出し、音韻情報を音韻継続時間長算出部 26 及びパラメータ接続部 27 に出力し、韻律情報をピッチパターン生成部 28 に出力する。

【0028】音韻継続時間長算出部 26 では、記号列解析部 25 からの音韻情報及び辞書 24 からの音韻継続時間長情報より各音韻の継続時間長を算出する。例えば、「白い花 (shiroihana)」というテキストが入力された場合、この「白い花」の 3 番目の母音（すなわち「i」）の継続時間長を求める場合を以下に説明する。

【0029】この母音「i」の継続時間長は（1）式より求めることができる。すなわち具体的には母音「i」の継続時間長＝『／i／の場合の係数＋前の音韻が／o／の場合の係数＋前々方の音韻が／r／の場合の係数＋後の音韻が／h／の場合の係数＋後々方の音韻が／a／の場合の係数＋前が促音でない場合の係数＋後が促音でない場合の係数＋長音でない場合の係数＋呼気段落のモーラ長「s」の場合の係数＋呼気段落中の場合の係数』＋文のモーラ長 N の場合の係数＋文中の場合の係数となる。

【0030】次にこの式の『』で囲んだ項を予め辞書 24 に保持してある値に置き換えて、母音／i／の継続時間長を算出する。すなわち母音「i」の継続時間長＝『辞書 24 より読み出した値』＋文のモーラ長 N の場合の係数＋文中の場合の係数となる。

【0031】このように予め辞書 24 に各音韻の継続時間長を求めるのに必要な音韻継続時間長情報を用意し、この音韻継続時間長情報を用いて各音韻の継続時間長を（1）式より算出すれば、異なる単語が同じ音韻環境にあつても、それぞれの単語に応じた自然な継続時間長を

算出することができる。ここで未知語には継続時間長情報は存在しないので、未知語の場合には従来の方法により音韻の継続時間長を算出する。

【0032】このようにして算出された各音韻の継続時間長はパラメータ接続部 27 及びピッチパターン生成部 28 に出力される。パラメータ接続部 27 では、記号列解析部 25 からの音韻情報と音韻継続時間長算出部 26 で算出された音韻の継続時間長とに基づいて、音素片データベース 29 から選択した音素片データを接続し、パラメータ列を生成する。生成されたパラメータ列は音声合成部 30 に出力される。

【0033】他方、ピッチパターン生成部 28 では、記号列解析部 25 からの韻律情報、辞書 24 からのアクセント指令値及び音韻継続時間長算出部 26 で算出された音韻の継続時間長に基づいて、ピッチパターンを生成する。

【0034】ここで例えばフレーズ指令及びアクセント指令がそれぞれ 1 つずつの場合の従来のピッチパターンの生成過程を図 2 に示す。上述のように従来のピッチパターン生成過程において、入力として与えられるパラメータは予め大量の文章を分析して求めてた値であり、所定の規則によつてそれぞれに適した値を選択し、この選択した値をそれぞれアクセント指令の大きさ A1、開始位置 t1、終了位置 t2、応答速度 β 1 に使用してピッチパターンを生成している。

【0035】他方この実施例においては、予め辞書 24 に各単語に固有のアクセント指令値を保持しておき、辞書 24 より所定の単語のアクセント指令の大きさ A2、開始位置 t3、終了位置 t4、応答速度 β 2 を読み出し、これらを用いて図 3 に示すようなピッチパターンを算出することにより、単語特有のピッチパターンを得る。

【0036】このように予め辞書 24 に各単語ごとの固有のアクセント指令値を用意し、このアクセント指令値を用いることによりピッチパターンを生成すれば、同じアクセント型で同じモラ数の単語が文章中の同じ位置に出現してもそれぞれの単語特有のアクセントパターンを得ることができるので、単語特有のピッチパターンを得ることができる。ここで未知語の場合には、音韻の継続時間長の算出の場合と同様に従来の方法によりアクセント指令値を算出する。

【0037】このようにして生成されたピッチパターンは音声合成部 30 に出力され、音声合成部 30 においてパラメータ列及びピッチパターンより音声波形を合成し、D/A 部 31 を介して合成音声信号を出力する。

【0038】以上の構成において、辞書 24 に音韻継続時間長情報とアクセント指令値とを予め保持しておく。テキスト入力部 22 に、辞書 24 に登録されている語が入力された場合、パラメータ接続部 27 においては、辞書 24 に保持されている音韻継続時間長情報を用いて算

出した音韻の継続時間長及び音韻情報に基づいて音素片データのパラメータ列が生成される。またピッチパターン生成部 28 においては、算出した音韻の継続時間長、音韻情報及び辞書 24 に保持されているアクセント指令値に基づいてピッチパターンが生成される。

【0039】このようにして生成したパラメータ列及びピッチパターンは各単語固有のパラメータ列及びピッチパターンになるので、これらのパラメータ列とピッチパターンとを合成して得られる合成音声は一段と人間の発声に近い合成音声となる。

【0040】以上の構成によれば、テキスト解析用の辞書 24 に、各単語のアクセント指令値と音韻継続時間長情報とを追加して保持する。辞書 24 に登録されている語が入力された場合には、音韻継続時間長情報を用いて算出した音韻の継続時間長に基づいてパラメータ列を生成し、この算出された音韻の継続時間長及び辞書 24 に保持されているアクセント指令値に基づいてピッチパターンを生成することにより、人間の発声に一段と近い合成音声を出力することができる。

【0041】なお上述の実施例においては、辞書 24 にアクセント指令値及び音韻継続時間長情報の両方を保持する場合について述べたが、本発明はこれに限らず、いずれか 1 つの情報だけを辞書 24 に保持するようにしてもよい。

【0042】

【発明の効果】上述のように本発明によれば、メモリ手段にアクセント指令値及び又は音韻の継続時間長情報を予め保持し、音韻情報又は音韻情報及び音韻継続時間長情報に基づいて算出した音韻の継続時間長を用いて音素片データのパラメータ列を生成し、算出した音韻の継続時間長及び韻律情報又は算出した音韻の継続時間長、韻律情報及びアクセント指令値に基づいてピッチパターンを生成して、これらのパラメータ列及びピッチパターンにより音声波形を合成することにより、人間の発声に一段と近い合成音声を出力することができる。

【図面の簡単な説明】

【図 1】本発明によるテキスト音声合成装置の一実施例の機能構成を示すブロック図である。

【図 2】従来のテキスト音声合成装置により算出されるピッチパターンを示す特性曲線である。

【図 3】実施例におけるテキスト音声合成装置により算出されるピッチパターンを示す特性曲線である。

【図 4】従来のテキスト音声合成装置の機能構成を示すブロック図である。

【符号の説明】

1、21……音声合成装置、2、22……テキスト入力部、3、23……文章解析部、4、24……辞書、5、25……記号列解析部、6、26……音韻継続時間長算出部、7、27……パラメータ接続部、8、28……ピッチパターン生成部、9、29……音素片データベース

ス、10、30……音声合成部、11、31……D/A 部。

【図1】

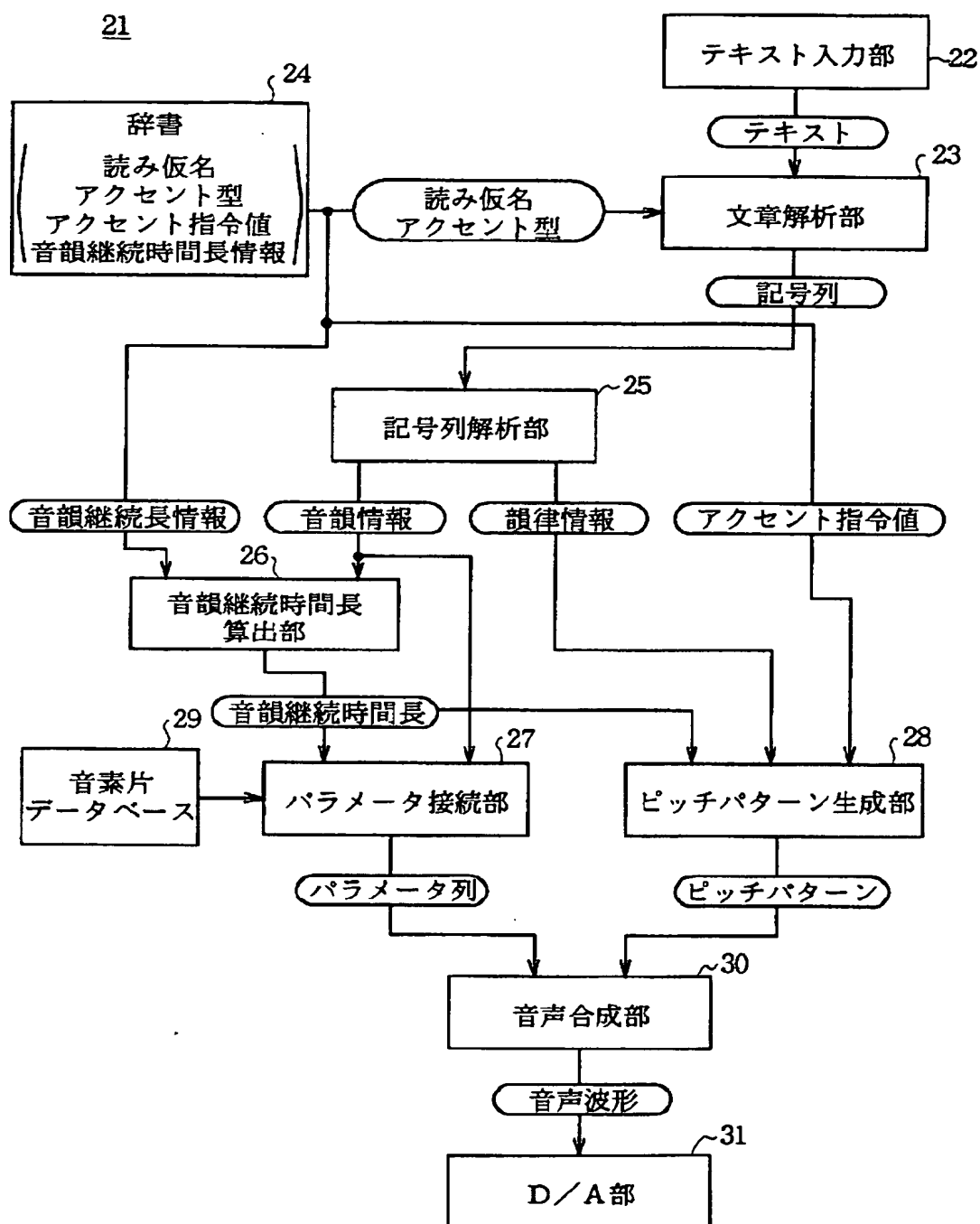


図1 実施例における音声合成装置の構成

【図2】

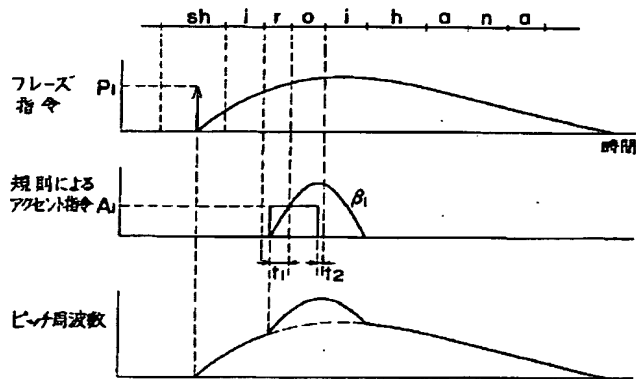


図2 従来の音声合成装置により算出されるピッチパターン

【図3】

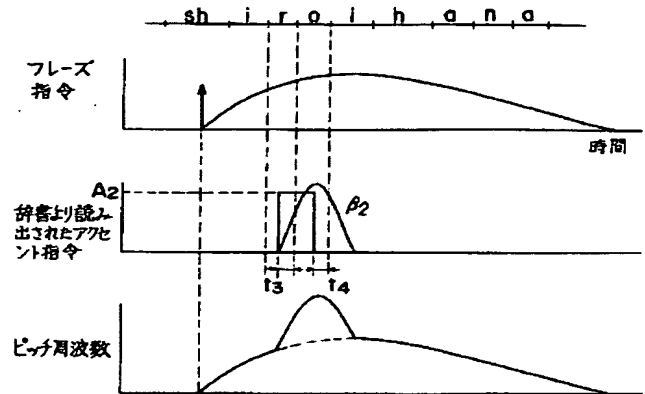


図3 実施例の音声合成装置により算出されるピッチパターン

【図 4】

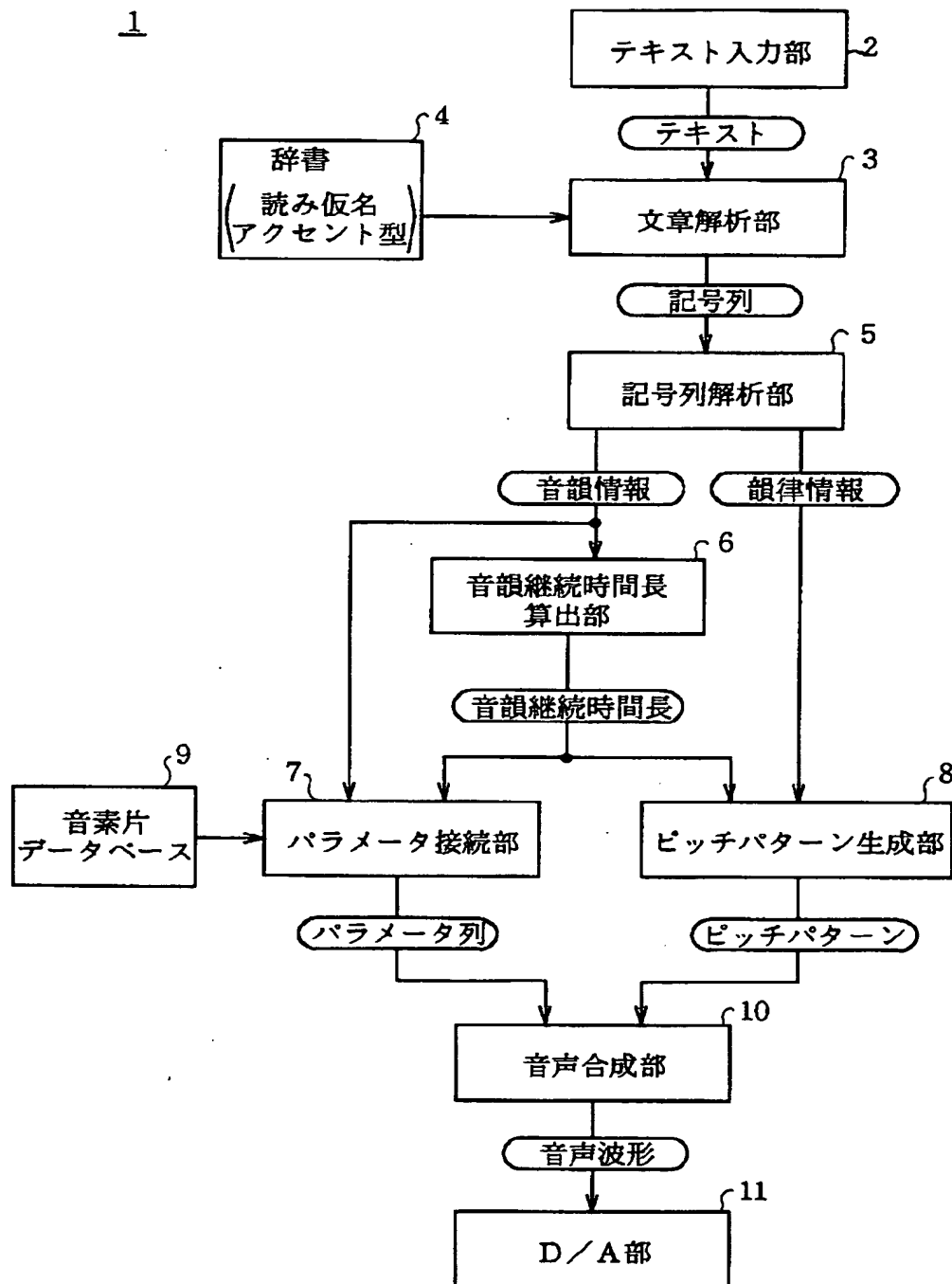


図 4 従来のテキスト音声合成装置の構成

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.